Share First, Ask Later (or Never?)
Studying Violations of GDPR’s Explicit Consent in Android Apps

Trung Tin Nguyen, Michael Backes, Ninja Marnau, and Ben Stock
CISPA Helmholtz Center for Information Security
General Data Protection Regulation (GDPR) governs all processing of personal data related to individuals situated in the EU and EEA.
In mobile apps, researchers have analyzed the app privacy policies to identify legislation violations, i.e., determining whether an app’s behavior is consistent with the privacy policy.

The GDPR knows several legal justifications for processing of personal data. In the case of transferring data to third party data controller for advertising purposes explicit consent is the viable option of these justifications.
The GDPR requires the consent to be *freely given, specific, informed*, and *unambiguous*.

Personal data transfer must only occur after the user has actively agreed (e.g., by clicking accept), i.e., “consent” packaged in terms and conditions or privacy policies is not compliant.
The community lacks insight into such GDPR violations in the mobile ecosystem.

Our research aims at answering the following research questions:

- **RQ1**: How many apps send out personal data without any prior consent?
- **RQ2**: Of the apps which send out any data, how many send it towards parties that act as data controllers under the GDPR?
- **RQ3**: Are developers aware of the requirements of GDPR and the issues that might arise from not following the outlined laws?
Overview of the methodology to identify violations of GDPR’s explicit consent in Android apps
Aiming to assess the state of GDPR violations in both high-profile and long-tail apps on the Play Store, and to understand if the violations are specific to either of them.

**High-profile app dataset**

16,163 top free high-profile apps from 33 app categories (i.e., AppBrain statistic).

**Long-tail app dataset**

70,000 distinct apps with at least 10,000 downloads and excluded those in the high-profile set.
Install the app and grant the necessary permissions

Run the app without any interactions

Intercept the TLS traffic by using MitM proxy
**PERSONAL DATA TIED TO A PHONE**

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAID</td>
<td>Android Advertising ID</td>
</tr>
<tr>
<td>BSSID</td>
<td>Router MAC addresses of nearby hotspots</td>
</tr>
<tr>
<td>Email</td>
<td>Email address of phone owner</td>
</tr>
<tr>
<td>GPS</td>
<td>User location</td>
</tr>
<tr>
<td>IMEI</td>
<td>Mobile phone equipment ID</td>
</tr>
<tr>
<td>IMSI</td>
<td>SIM card ID</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAC</td>
<td>MAC address of WiFi interface</td>
</tr>
<tr>
<td>PHONE</td>
<td>Mobile phone’s number</td>
</tr>
<tr>
<td>SIM_SERIAL</td>
<td>SIM card ID</td>
</tr>
<tr>
<td>SERIAL</td>
<td>Phone hardware ID (serial number)</td>
</tr>
<tr>
<td>SSID</td>
<td>Router SSIDs of nearby hotspots</td>
</tr>
<tr>
<td>GSF ID</td>
<td>Google Services Framework ID</td>
</tr>
</tbody>
</table>
Using simple string-matching to identify personal data that is known, common transformations such as upper/lower case, hashing (e.g., MD5), encoding (e.g., base64) are considered.

POST https://api.uca.cloud.unity3d.com HTTP/1.1

{"header": {"appid": "323c504d-fae5-449d-acd1-a89f2cf06b09", "userid": "8190a000-0b24-4f36-a981-c535f57ff164", "sessionid": 11219588307516230, "platform": "Android", "sdk_ver": "u5.3.3f1"}, "events": [{"type": "deviceInfo", "ts": 1607086918599, "make": "Android", "model": "Google Pixel 3a", "processor_type": "ARMv7 VFPv3 NEON", "system_memory_size": "3593", "engine_ver": "5.3.3f1", "app_name": "com.xxxx", "app_install_mode": "dev_release", "debug_build": false, "license_type": "personal", "os_ver": "Android OS 9 / API-28 (PQ3B.190801.002/5674421)", "deviceid": "cf9f2bb31b46f4871094b3217b8349a9", "app_ver": "1.0.21", "changed": [{"app_ver", "os_ver", "sdk_ver"}], {...}, {"type": "deviceInfo", "ts": 1607086918086, "adsid": "70831fd5-c2df-4b75-94bd-915a2046fe14", "ads_tracking": false, "changed": ["adsid"]}}

Advertising ID: 70831fd5-c2df-4b75-94bd-915a2046fe14
Using simple string-matching to identify personal data that is **known**, common transformations such as upper/lower case, hashing (e.g., MD5), encoding (e.g., base64) are considered.

**IMEI:**
354787113965960

**MD5 Hashing:**
cf9f2bb31b46f4871094b3217b8349a9

**POST https://api.uca.cloud.unity3d.com HTTP/1.1**

```
{"header": {"appid": "323c504d-fae5-449d-acd1-a89f2cf06b09", "userid": "8190a000-0b24-4f36-a981-c535f57ff164", "sessionid": 11219588307516230, "platform": "Android", "sdk_ver": "u5.3.3f1"}, "events": [{"type": "deviceInfo", "ts": 1607086918599, "make": "Android", "model": "Google Pixel 3a", "processor_type": "ARMv7 VFPv3 NEON", "system_memory_size": "3593", "engine_ver": "5.3.3f1", "app_name": "com.xxxx", "app_install_mode": "dev_release", "debug_build": false, "license_type": "personal", "os_ver": "Android OS 9 / API-28 (PQ3B.190801.002/5674421)", "deviceid": "cf9f2bb31b46f4871094b3217b8349a9", "app_ver": "1.0.21", "changed": ["app_ver", "os_ver", "sdk_ver"]}, {...}, {"type": "deviceInfo", "ts": 1607086919086, "adsid": "70831fd5-c2df-4b75-94bd-915a2046fe14", "ads_tracking": false, "changed": ["adsid"]}}
```
Performing multiple runs with a different set of devices to identify parameters that could be used to track and profile an individual, but do not obviously string-match

<table>
<thead>
<tr>
<th>Domains</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>appsflyer.com</td>
<td>deviceFingerPrintId=&lt;UUID&gt;</td>
</tr>
<tr>
<td>branch.io</td>
<td>hardware_id=6fd9a2e0f2721498</td>
</tr>
<tr>
<td>tapjoy.com</td>
<td>managed_device_id=tjid.36cec2b4196...</td>
</tr>
<tr>
<td>unity3d.com</td>
<td>common.deviceid=d3d55baf21d8f31839...</td>
</tr>
</tbody>
</table>
Third-party domains representing only 12.0% of domains which received personal data, are responsible for 94.7% of cases of receiving personal data without prior consent.

86,163 apps

72,274 (83.9%) apps were successfully analyzed

28,665 (39.66%) sent personal data to 1,744 domain names (209 third-party domains)
An app which relies on external data controllers for targeted advertising needs to explicitly ask for the user’s consent to share her personal data with the third party.
24,838 (88.5% 28,065) apps sent personal data to ad-related domains, thereby violating GDPR’s mandated consent.

- More than half of the apps which sent data without consent sent data to (at least) Facebook, then Unity, and Flurry.
- Library providers make it very cumbersome for developers to be compliant with GDPR.

3,840 apps that combined the AAID with some other type of personal information (e.g., IMEI).
- Due to developers’ opt-in or the usage of outdated libraries that do not support GDPR.

The phenomenon of sending out personal data without prior explicit consent happens as frequently and with as many parties in both dataset.
- 11,914 developers were notified (responsible for 17,795 apps).
- Until February 1, 2021: there are 2,083 apps accessed the notification reports.
  - 448 distinct developers that answered the survey.

GDPR issues are widespread, often misunderstood, and require effort from advertisement providers, app stores, and developers alike to mitigate the problems.
CALL TO ACTIONS

■ Third Parties Should Take Responsibility
  ○ Limiting the data collection.
  ○ Providing the automatically consent mechanism.
  ○ Making their documentation transparent and easy to access.

■ App Stores Should Take Actions
  ○ Employing such techniques as our to identify the potential violations of GDPR explicit consent, or the usage of outdated SDK.

■ Support for Developers
  ○ Strongly call on third-party vendors for better documentation and transparency in legal documents.
CONCLUSION

MOTIVATION AND RESEARCH QUESTIONS

The community lacks insight into such GDPR violations in the mobile ecosystem. Our research aims at answering the following research questions:

- **RQ1:** How many apps send out personal data without any prior consent?
- **RQ2:** Of the apps which send out any data, how many send it towards parties that act as data controllers under the GDPR?
- **RQ3:** Are developers aware of the requirements of GDPR and the issues that might arise from not following the outlined laws?

METHODOLOGY

Overview of the methodology to identify violations of GDPR’s explicit consent in Android apps

- **App Dataset Construction**
- **Network Traffic Analysis**
- **Identifying Advertising Domains**
- **Developer Notification**

IN-DEPTH ANALYSIS OF VIOLATIONS

- **24,838** (88.5% 28,065) apps sent personal data to ad-related domains, thereby violating GDPR’s mandated consent.
  - Library providers make it very cumbersome for developers to be compliant with GDPR.
- **3,840** apps that combined the AAIO with some other type of personal information.
  - Due to developers’ opt-in or the usage of outdated libraries that do not support GDPR.
- The phenomenon of sending out personal data without prior explicit consent happens as frequently and with as many parties in both datasets.

DEVELOPER NOTIFICATIONS

- **11,914** developers were notified (responsible for 17,795 apps).
- Until February 1, 2021: there are 2,083 apps accessed the notification report.
  - 448 distinct developers that answered the survey.

CALL TO ACTIONS

- **Third Parties Should Take Responsibility**
  - Limiting the data collection.
  - Providing the automatically consent mechanism.
  - Making their documentation transparent and easy to access.
- **App Stores Should Take Actions**
  - Employing such techniques as ours to identify the potential violations of GDPR explicit consent, or the usage of outdated SDK.
- **Support for Developers**
  - Strongly call on third-party vendors for better documentation and transparency in legal documents.

GDPR issues are widespread, often misunderstood, and require effort from advertisement providers, app stores, and developers alike to mitigate the problems.
Thank you!
tin.nguyen@cispa.de